IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 22, 2017, accepted May 9, 2017, date of publication May 23, 2017, date of current version June 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2707394

Photo-to-Sketch Transformation in a Complex Background

XIANLIN ZHANG[1], XUEMING LI[2], SHUXIN OUYANG[1], AND YANG LIU[3]

[1]Institute of Information and Communication, Beijing University of Posts and Telecommunications, Beijing 100089, China
[2]Beijing Key Laboratory of Network System and Network Culture, Institute of Digital Media and Design, Beijing University of Posts and Telecommunications, Beijing 100089, China
[3]Institute of Digital Media and Design, Beijing University of Posts and Telecommunications, Beijing 100089, China

Corresponding author: Xianlin Zhang (zxlin@bupt.edu.cn)

ABSTRACT We investigate the problem of sketch generation for sketch-based image retrieval (SBIR). Solving this problem is important to obtain better retrieval results, because a powerful feature extraction algorithm is inefficient. Transforming photos from raw pixels into pseudo-sketches closes the gap between these two domains and plays a significant role in SBIR. This problem is relatively challenging because of: 1) the complexity of the image background and 2) the complexity of the resulting edge map. In this paper, we develop a system to generate pseudo-sketches from photos. Saliency detection is used to extract the major objects from a photo. Then, a Gabor filter is designed to further capture the ''real-major'' object. Finally, the Sobel operator is used to obtain the final pseudo-sketch. Experiments are conducted using the Flickr15k data set. The results show that the obtained pseudo-sketches are reasonable and that the SBIR process produces the state-of-the-art results in certain categories.

INDEX TERMS Gabor filter, pseudo-sketches, saliency detection, SBIR, Sobel operator.

I. INTRODUCTION

With the development of digital devices, sketch-based image retrieval (SBIR) [1] has recently received a large amount of attention in the computer vision (CV) field. This process is different from text-based retrieval because visual sketch queries are intuitive when depicting the shape or appearance of an object. ''Sketches'' may be produced by either professional artist (such as frontal face sketch images with rich facial structure details) or non-professionals (such as free-hand sketches generated using electronic touch-screen devices). Furthermore, several works exist on face sketch synthesis [2] based on recognition. Wang et al. presented a transductive face sketch-photo synthesis method [3] that incorporates test samples during learning to achieve optimal performance, and they also recently proposed a Bayesian framework [4] for face sketch synthesis.

In this paper, a ''sketch'' is defined as a free-hand sketch produced by an amateur using a touch screen device (e.g., an iPad, a mobile phone or a Wacom tablet). In this case, a key challenge arises in SBIR, which is how to overcome the ambiguity inherent in a sketch [5] since a query sketch can exhibit a large amount of variation (such as transformation, scaling, and rotation). Moreover, sketches that are produced by different users, even for the same object, can appear unrefined or substantially different due to poor drawing skills or a limited time for drawing. Recent studies on SBIR can be divided into two aspects: how to make a better feature operator and how to generate a better sketch from a natural image during the procedure of SBIR. To date, there has been little work focused on the latter step.

Our work focuses on obtaining a concise and simple sketch from a single image. The main contributions are the following: (1) For the first time, an improved random walk model based on a Markov chain is used to obtain saliency maps from complex background images for SBIR. (2) A specific Gabor filter is designed to refine the saliency maps for pseudo-sketch generation. (3) Experimental results demonstrate that with the pseudo-sketch, the SBIR accuracies are improved, particularly in certain categories, in which state-of-the-art results are achieved.

The content of this paper is organized as follows. Section II describes the related work. In Section III, the algorithm is illustrated in detail. Experiment settings and results are presented in Section IV. Conclusions are drawn in Section V.

II. RELATED WORK

The majority of feature extraction methods have been based on the histogram of orientations approach, and sketches are typically represented by global or local descriptors. The histogram of oriented gradient (HOG) [6] descriptor

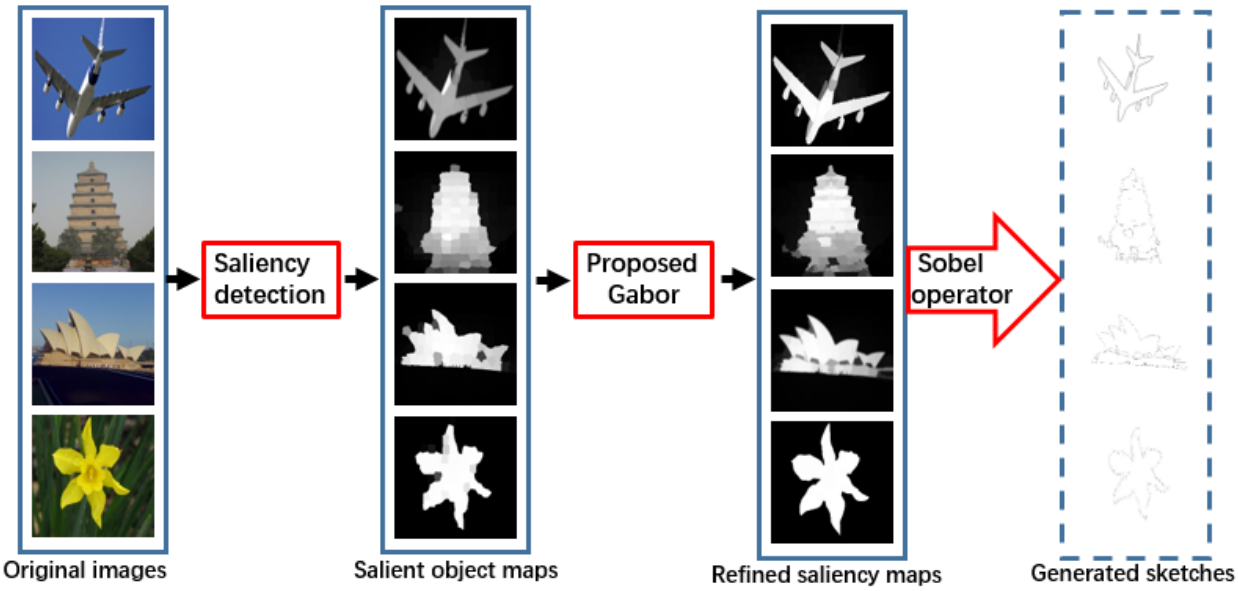


FIGURE 1. Overview of the complete processing approach.

has been impactful in SBIR development. Eitz proposed the bag-of-words (BOW) model [7] based on the HOG. K-Chatfield proposed an approach [8] that uses the local self-similarity features of an image in combination with the BOW model to perform subsequent processing for sketch retrieval. Saavedra et al. proposed a histogram of edge local orientations (HELO) [9]. Based on this work, Jose M Saavedra proposed the soft-HELO (S-HELO) [10] descriptor, which was inspired by both HELO and HOG. These descriptors have all proven to be effective for SBIR.

Almost all SBIR algorithm research has focused on developing techniques for extracting features [11]–[15] to improve the retrieval accuracy. Nevertheless, these studies all ignore to some extent the influence of the generated sketch in the previous stage. Most of the work uses only the simple Canny edge detector to obtain the corresponding sketches from photos. Recently, a sketch generation algorithm was developed using the methods of contour detection and object segmentation [16]–[18]. Ablaze proposed a type of edge detector to obtain the hierarchical segmentation region, intending to use the edge that was produced by the optimal segmentation as the final sketch. Marvaniya *et al.* [19] proposed a sketch generation algorithm to produce sketches of the same image type. Guo *et al.* [20] proposed a primal sketch algorithm that uses a single image to obtain the sketch; it combines two types of generative models, namely, sparse encoding and Markov random field. Qi *et al.* [21] proposed a simpler way to generate a sketch; this method is based on perceptual grouping combined with multiple gestalt rules. His method can obtain a sketch from a single image; however, the results lack simplicity and clarity.

This paper focuses on a sketch generation technique that significantly improves the results of produced sketches. Saliency detection based on a Markov absorption model is used to obtain the salient region map. To reduce the noise and retain the orientation of the edge, a specific Gabor filter is proposed to refine the edge map. The Sobel operator is used to abstract the final similar form of the sketches. Fig. 1 shows the complete process flowchart.

III. THE PROPOSED ALGORITHM

A. THE ABSORBING MARKOV CHAIN MODEL

We can reduce the complexity of image post-processing through the use of a small number of superpixels (instead of all the image pixels) to express the picture features. First, we apply a superpixel segmentation algorithm to complete the transformation from the original image to the superpixels. There are many ways to finish this step in practice, and an existing technique [22] is used to obtain high validity. Then, a graph is represented by $\Lambda = (U, \Gamma)$, where U denotes the superpixels and Γ denotes a series of non-directional edges that represent the connectivity between two different superpixels. We connect two vertically adjacent superpixels if they have the same values. The edge that combines two nodes i and j is shown as n_{ij}, while h_{ij} expresses the weight assigned to the edge n_{ij}. Finally, h_{ij} is given as

$$h_{i,j} = exp\left\{-\left\|l_i - l_j\right\|^2 \left(2\sigma^2\right)^{-1}\right\} \tag{1}$$

where the features l_i and l_j correspond to the two codes i and j, respectively, and σ is a weight constant. The affinity matrix of the image graph model can be written as $H = \left(h_{i,j}\right)_{mm}$. (The basic concepts of the Markov chain are described elsewhere [23]–[26].) Based on the graph model, [27] duplicated the top and left boundary superpixels as absorbing nodes of the Markov chain, and the remaining nodes in the graph

were taken as transient states. That study used the model $\Lambda' = (U', \Gamma')$, which includes the virtual nodes to form the final absorbing chain. Here, $Z = (z_{ij})_{mk}$ is a different affinity matrix, z_{ij} is the weight between node i and the virtual absorbing node j in the image, and the probability translation matrix is ultimately formed as follows:

$$\Pi = \begin{bmatrix} K_1 & Y_1 \\ O & I_1 \end{bmatrix} \tag{2}$$

where O is the k^*m zero matrix, I_1 is the k^*k identity matrix, $K_1 = J^{-1}H$, $Y_1 = J^{-1}Z$, and Q_1 is the transition probability matrix. J is the sum of the matrixes $diag\left\{j_{11}^H, j_{22}^H, \ldots j_{mm}^H\right\}$ and $diag\left\{j_{11}^Z, j_{22}^Z, \ldots j_{mm}^Z\right\}$. The final absorption probability matrix is defined as

$$\Phi_1 = M_1 Y_1 \tag{3}$$

$$M_1 = (I - Q_1^{-1}) = I + Q_1 + Q_1^2 + \ldots \tag{4}$$

B. THE IMPROVED MARKOV MODEL TO OBTAIN THE SALIENCY MAP

Based on the knowledge of the transient state and the absorbing probability in the Markov chain, [27] assumed the partially duplicated image boundaries to be absorbing nodes and the remaining nodes in the image to be transient states. However, this approach could cause over-inhibition and make the salient region less easily detectable. Nevertheless, we employ all the boundary nodes as absorbing nodes since this approach will exploit the graph information comprehensively and satisfactorily cover the graph under the entire ergodic Markov chain. This procedure is different from that described in the reference. A portion of the results is shown in Fig. 2. The original method in the first column excessively inhibits the salient region, which will directly influence the final edge map. The results shown in the improved model in column three avoid this problem for certain images with a cluttered background.

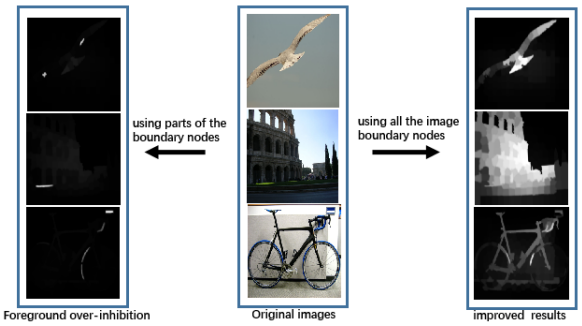


FIGURE 2. Comparison of image results obtained using the improved method.

The algorithm is as follows. We can obtain the corresponding translation matrix $\overline{\Pi}$ and the absorbing probability matrix $\overline{\Phi}$ through formulas (2)-(4) using all image boundary nodes. The probability $\overline{\varphi_{ij}}$ denotes the similarity between the superpixel node i and all the absorbing image nodes j. Since the absorption probability matrix $\overline{\Phi}$ has been obtained in the new way and is known, we rearrange the absorption probability values of each node i with all the other nodes j in the image in descending order:

$$\overline{\varphi'_{i,1}} \geq \overline{\varphi'_{i,2}} \geq \ldots \geq \overline{\varphi'_{i,k}}, \quad \left(\overline{\varphi'_{i,j}} \in U_{j=1}^{k} \overline{\varphi_{i,j}}\right) \tag{5}$$

We construct the new matrix $\overline{\Phi'_{m,k}}$ and then select the first r columns, which show strong relevance to node i of matrix $\overline{\Phi'_{m,k}}$. Therefore, another new matrix is also produced, namely, $\overline{\Phi'_{r,k}}$. The saliency of node i is defined as

$$M_\varphi(i) = \frac{1}{2}\left(\begin{array}{l} exp\left(M_{\varphi up}(i)\right) + M_{\varphi lf}(i) \\ +M_{\varphi lr}(i) + M_{\varphi lld}(i) \end{array}\right) \tag{6}$$

$$M_\varphi(i) = exp(1 - bm(i)).zm(i), \quad bm(i) = \sum_{j=1}^{r} \overline{\varphi'_{ij}} \tag{7}$$

This finding shows the similarity to the image background, and $zm(i) = \sum_{j=r}^{k} \overline{\varphi'_{ij}}$ reveals the dissimilarity to the backdrop. While the top nodes are approximated to the background, we emphasize the background through an exponential function. A crude edge map is realized by suppressing the background.

Following the background suppression, the superpixels that correspond to the salient region and the foreground nodes can be detected. We assume that the foreground nodes are absorbing nodes and view the others as transient nodes, in turn, to finish addressing the foreground's outstanding aspects. Similar to the matrix Y_1, the matrix Y^*, which considers all the boundary nodes, is obtained:

$$Y^* = \begin{bmatrix} y_{11}^* \ldots . y_{1l}^* \\ \ldots\ldots\ldots \\ y_{m1}^* \ldots . y_{1l}^* \end{bmatrix} \tag{8}$$

A new matrix M_{mm}^* is constructed, as discussed in Section III:

$$M^* = I^* + Q^* + \left(Q^*\right)^2 + \left(Q^*\right)^3 + \ldots \tag{9}$$

We multiply the matrixes M^* and Y^*, the result of which is written as $= M^*Y^*$, and we rearrange each row of matrix Φ^* in descending order.

$$\varphi_{i1}^* \geq \varphi_{i2}^* \geq \ldots \geq \varphi_{il}^*, \quad \left(\varphi_{ij}^* \in U_{j=1}^{l} \varphi_{ij}\right) \tag{10}$$

Here, φ_{ij}^* can be interpreted as the similarity of node i to the foreground node j. We sum up the top g numbers of each row φ_{ij}^* to represent the similarity of the image node i to the foreground of an image. Then, the new saliency of each node i is defined as

$$M_f(i) = \sum_{j=1}^{g} \varphi_{ij}^* \tag{11}$$

The saliency map is roughly detected at that time. Subsequently, we optimize the edge map by background suppression and foreground improvement. Inspired by [28], we first refine the saliency edge map by a mathematical mechanism that effectively highlights the object region. A clustering algorithm is applied to group the nodes of the input map

into U clusters; for node n_i, $\kappa_i = \{n_j\}_{j=1}^{m}$ presents the remaining nodes in the same cluster. We refine the image map using interpolation of the nodes that are in the same cluster within it. The node n_i with its new saliency is defined as in (12), and we set χ equal to 0.4 in the experiments. The total sum of variance of the Lab space in each feature dimension is σ_L.

$$T_m^*(n_i) = \chi T_m(n_i) + (1-\chi)\sum_{j=1}^{m} F_{ij} T_m(n_i) \left(\sum_{j=1}^{m} F_{ij}\right)^{-1} \tag{12}$$

$$F_{ij} = -exp\left\{\left\|k_i - k_j\right\|^2 (2\sigma_L)^{-1}\right\} \tag{13}$$

After the highlight processing measure, it is crucial to eradicate additional noise on the map. To solve this problem, we then apply a depression function to weaken the background. The function is defined as in (14); the background, in which there is noise and unnecessary detail on the image edge map, is suppressed satisfactorily through the depression formula.

$$V(x) = x * (2x), \quad x \le 0.5 \tag{14}$$

C. THE SPECIFIC PROPOSED GABOR REFINEMENT STRATEGY

To maintain the well-formedness of the edges and to better optimize the saliency map, a further refined step is performed. To retain robustness in terms of rotation invariance and preserve the local feature details in the image map, a two-dimensional Gabor filter [29]–[32] is applied because it can maintain better local structural information on the spatial frequency (scale), spatial position and orientation selectivity. In addition, the Gabor filter is a linear filter for edge detection in both the frequency and spatial domains, and a two-dimensional Gabor filter is a product of a sinusoidal plane wave and a Gaussian kernel function. The filter function is defined as

$$g(x,y) = \frac{1}{2\pi\delta_u\delta_v} exp\left\{-\frac{1}{2}\left(\frac{u^2}{\delta_u^2} + \frac{v^2}{\delta_v^2}\right)\right\} cos(\omega u) \tag{15}$$

$$u = xcos\theta + ysin\theta; \quad v = xcos\theta - ysin\theta \tag{16}$$

where θ is the direction of the filter and δ_u and δ_v are the standard variances of the coordinate axes u and v, respectively. Here, ω, which defines the Gabor filter's own selective characteristics, is the frequency of the sinusoidal function.

As shown in Fig. 3, we obtain enhanced image saliency maps using a specific Gabor filter. The refined butterfly edge map is relatively simple and clear. From the local enlarged images in the red rectangle, it is easy to distinguish that the refined map is better in terms of the edge details than the previous map. The selection parameters of the Gabor filter are important in practice. Different parameters can produce a large variation for the same saliency map. In the experiments, we set the orientation of the Gabor kernel function to be equal to 8, the frequency that determines the wavelength of the Gabor to 8, and the standard variance to be equal to $2^2 * pi$ (the initial phase is not considered); this set of parameters has performed well for most of the saliency maps and is better at both retaining the edges and improving the overall appearance of the maps.

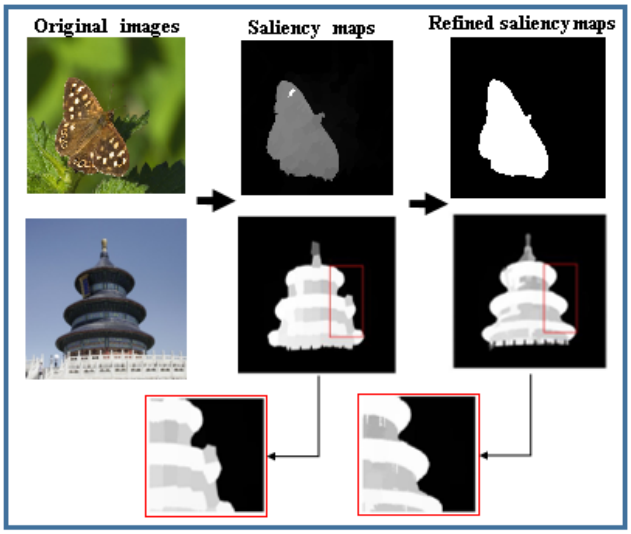


FIGURE 3. Saliency maps refined with a specific Gabor filter. The images are from the categories of butterfly and Beijing Temple of Heaven.

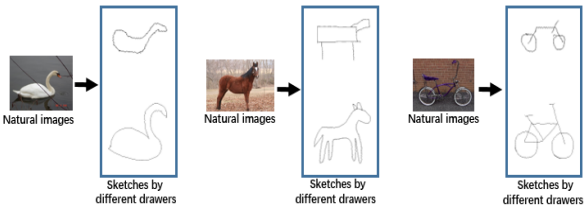


FIGURE 4. Examples of sketches produced by different drawers.

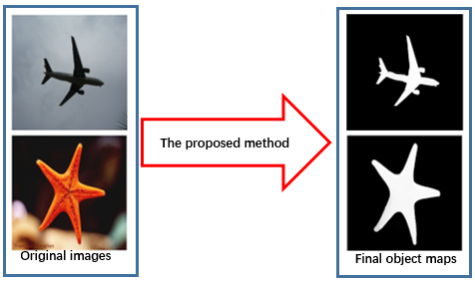


FIGURE 5. Examples of contour maps produced by the proposed method.

In Section III, tasks to highlight the foreground and eliminate the dispensable background on the edge maps have been performed effectively using two mathematical techniques. To further improve the results, the saliency map is filtered by a proposed specific two-dimensional Gabor filter. The final saliency map is then obtained. To abstract the sketch from the saliency map, the Sobel operator is applied last. The entire process is summarized in algorithm 1.

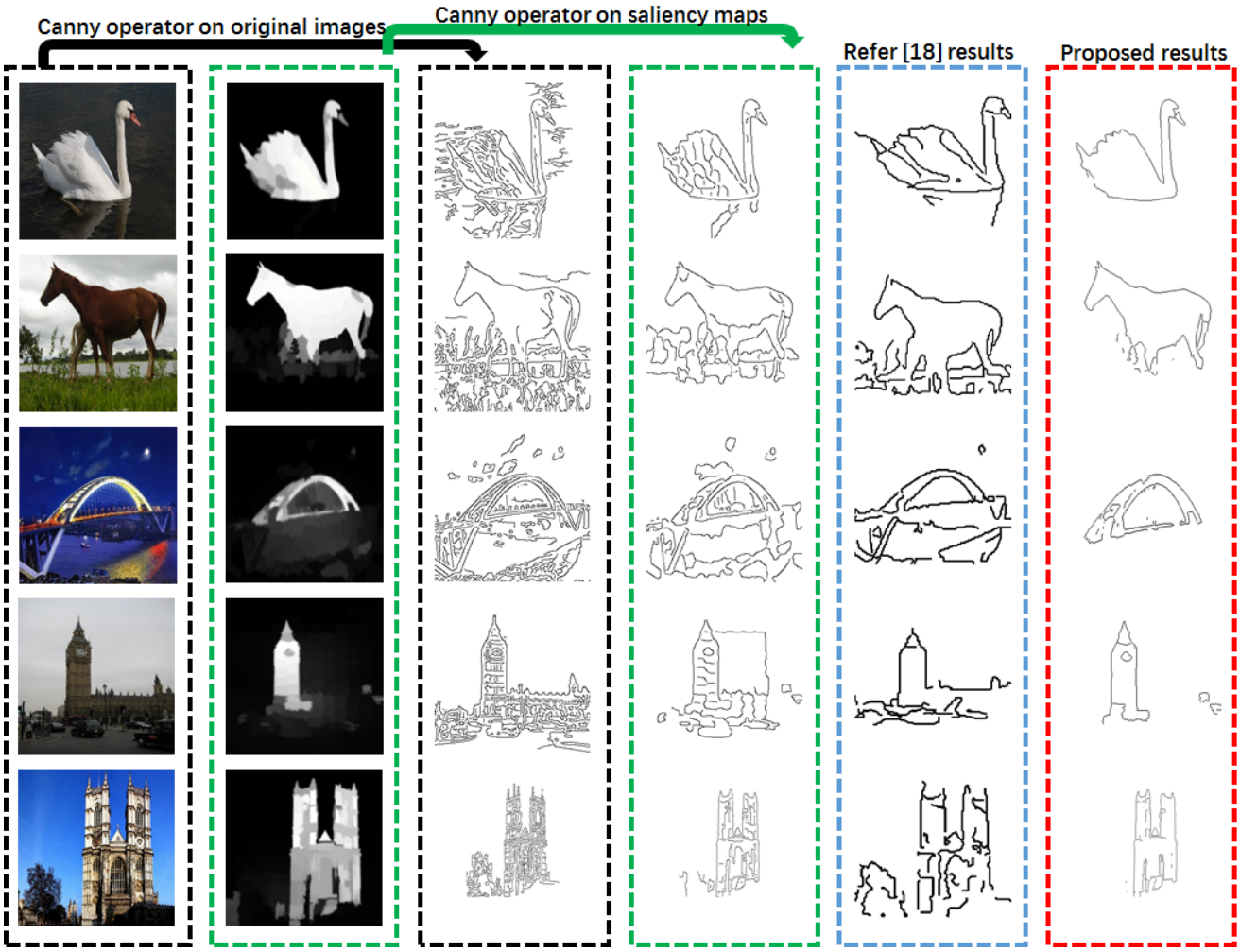


FIGURE 6. Comparisons of sketches generated by different algorithms.

In addition, the experimental results described in the following section will prove that our method is impactful for SBIR; in particular, they are superior in certain specific image categories.

Algorithm 1 Sketch Generation

Input: A random image x from a category in the database
Step 1: Convert the image into superpixels and represent it in the form of a graph Λ
Step 2: Construct the translation matrix $\overline{\Pi}$ and the absorption probability matrix $\overline{\Phi}$ and then obtain the saliency map through background inhibition and foreground highlighting
Step 3: Refine the measures to obtain the saliency map V
Step 4: Apply a specific Gabor filter g to obtain the final map G
Step 5: Apply the Sobel operator to the final edge map
Output: The generated pseudo-sketch S of the input image

IV. EXPERIMENTAL EVALUATION

A. EXPERIMENTAL DATASET

In this paper, we evaluate our algorithm on the Flickr15k dataset. The dataset contains approximately 15,000 images divided into 33 categories, and iconic classes such as airplane, duck and horse are all included. The use of this dataset is challenging because the number of images in each category varies greatly, from 21 images (horse class) to 1600 images (Eiffel Tower class) and because the background complexities of the images are not uniform. Furthermore, the query sets of the dataset consist of 10 folders, and each folder contains 33 sketches that correspond to the 33 categories mentioned above, while 10 non-expert sketchers are added to each folder to provide free-hand sketch queries. The query sketches, which lack the refinement associated with professional drawing skills, also have large differences. Fig. 4 shows examples of the query sketches.

In fact, some of the images in Flickr15k have no corresponding ground-truth labels; thus, we refresh the images and their corresponding ground-truth labels. Finally, we obtain a

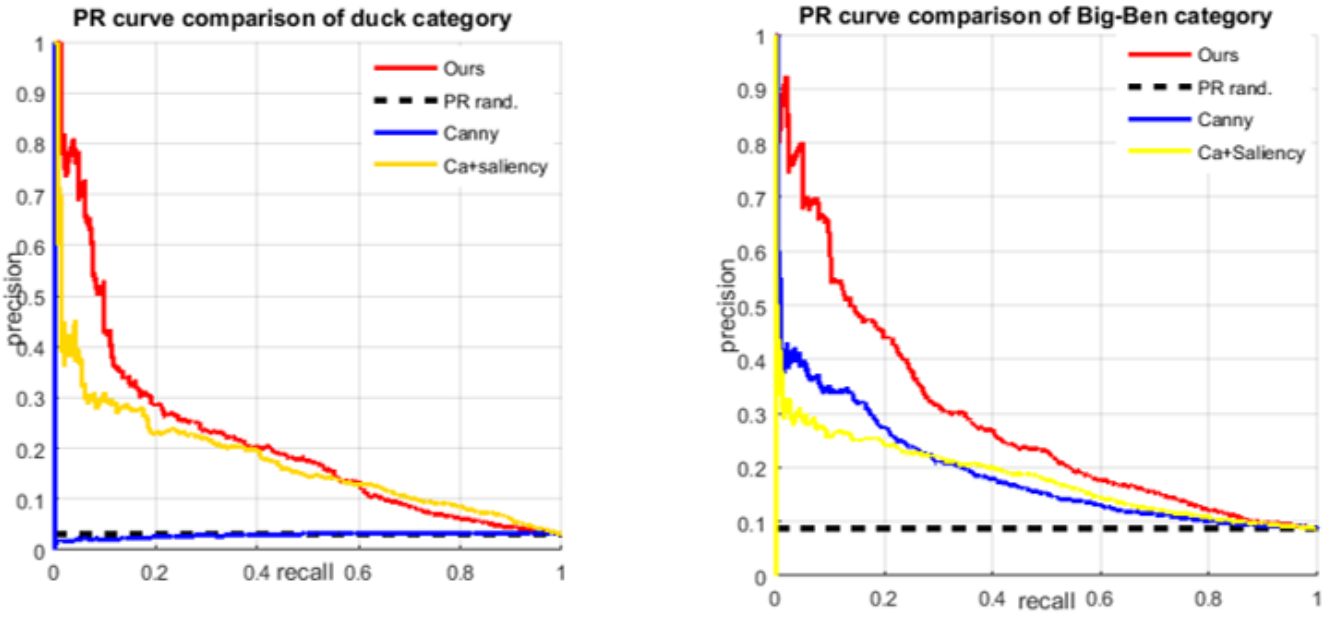


FIGURE 7. PR curve comparisons for the categories of duck and Big Ben.

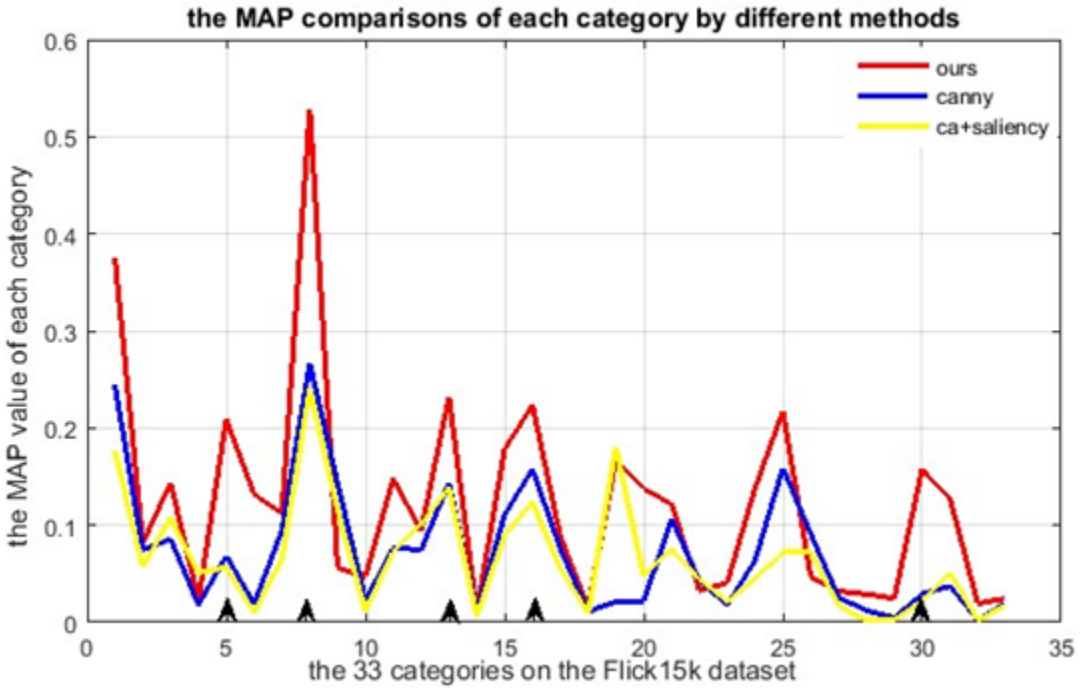


FIGURE 8. MAP comparisons for all 33 categories processed using different methods.

total number of 14502 images, each of which has a unique ground-truth label. The experimental environment consists of a Dell PowerEdge M610 blade server and an Acer notebook with 6 GB of memory, and all simulations are performed using MATLAB.

B. RETRIEVAL EXPERIMENTS ON FLICKR15K

Since the method of saliency region detection is adopted to remove the image background and to produce a better and concise form of the sketches, we now show the overall results. Fig. 5 shows selected results. The saliency maps are encouraging.

The Sobel operator is a discrete first-order differentiable operator that offers high efficiency for edge detection in image processing. To obtain an ultimately similar form of the sketches, the Sobel operator is applied to the saliency maps. In the experiments, we compare the produced pseudo-sketches using several baselines, and selected results are shown in Fig. 6.

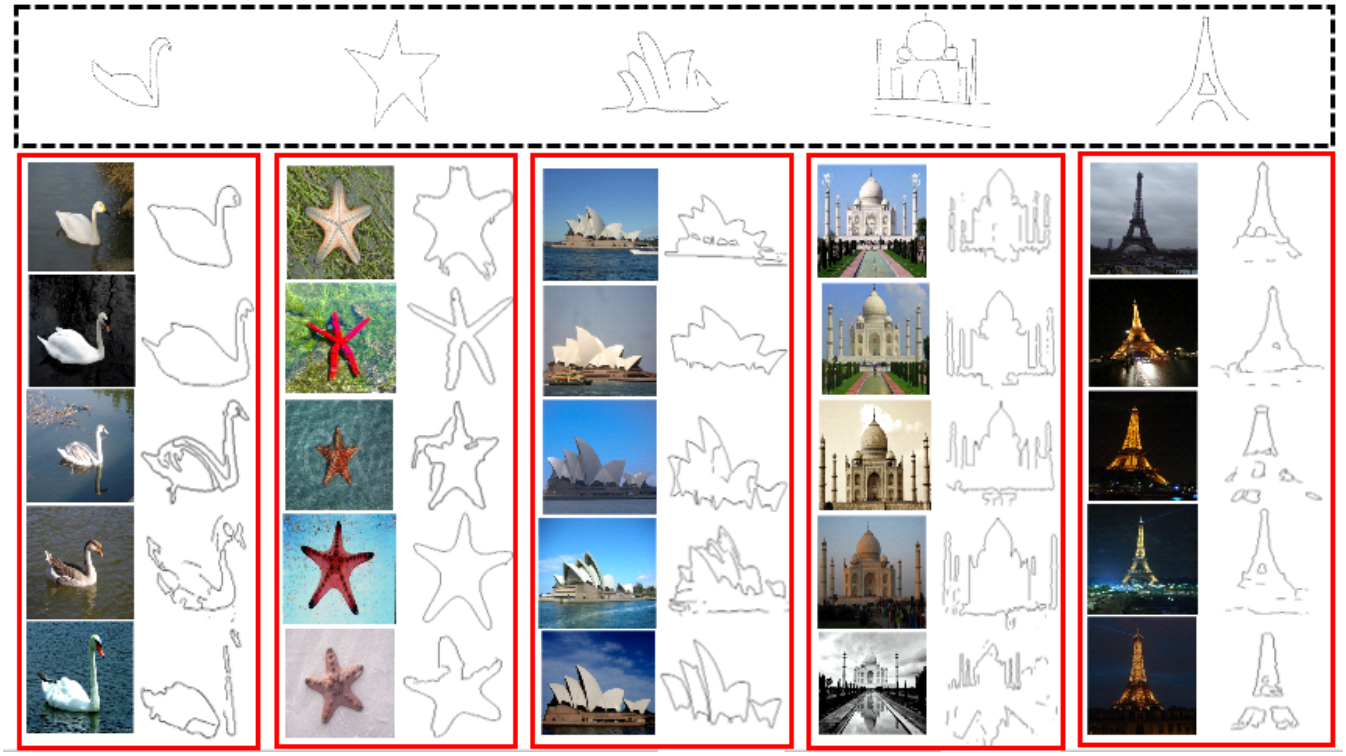


FIGURE 9. Parts of the query sketches and retrieval results from the Flickr15k dataset.

The results reveal that the sketches generated using the proposed saliency method are better than the other sketches. This finding establishes the suitability of the proposed approach.

Using the Canny method, the edges of the original image emerge on both the object region and the unused background, as shown in the third column in Fig. 6. In fact, because the query sketches are produced by non-expert drawers, most of the sketches are of amateur quality. Thus, the use of the Canny method for obtaining the edges of the image could be regarded as ineffective because it yields edges of the entire image, which may affect the subsequent processing procedure and eventually influence the final retrieval efficiency. In the fourth column, sketches are generated by applying the Canny method on the saliency map. The results are not ideal because a large amount of extra noise is included. This outcome will have an impact on feature extraction and similarity matching. The outcomes that arise from the fifth column are more satisfactory in terms of their appearance, but non-related component noise, which is undesirable in applications, still exists. This phenomenon is prominent in the Westminster Abbey image in the last row of Fig. 6. In contrast, sketches generated by the improved algorithm in the last column include the total object region and contain little noise. The generated sketches are more concise and well-formed. The following retrieval experiments prove the method to be useful.

By comparing the average precision (AP) on specific image categories, the efficiency of the proposed method is proven. Moreover, the PR curves obtained by different methods applied to the duck and Big Ben categories are shown in Fig. 7.

In the duck image category (in which the total number of images is 454), the maximum AP achieved by directly using the Canny method is approximately 3.3%, and the Canny method based on the saliency map can reach a value of 16.68%. However, with our method, the AP is 22.84%. In the Big Ben image category (in which the total number of images is 1267), the Canny method applied with the saliency approach produces an AP of 17.91% and the Canny method alone achieves a value of 18.87%, but the AP value of our proposed technique is as high as 32.7%, which is better than those of the other methods. Our novel method delivers superior results relative to the other techniques. To compare the MAP value of each category by different algorithms in a clear form, a chart presenting all 33 categories is shown in Fig. 8.

The experimental performance indicates that for the category of duck, Big Ben, and so on, which are indicated by the black arrows in the sequence, our novel method performs much better than the other techniques. It is clear that the value corresponding to the Canny operator is relatively low. The Canny method based on the saliency measure does not perform well in certain categories. The novel method can obtain a higher MAP value for most of the categories, such as the eighth category (corresponding to 1606 images of the Eiffel Tower) and the seventeenth category (corresponding to 184 images of the Taj Mahal).

In summary, the proposed method can provide a strong MAP value for most of the categories. For specific image categories with cluttered backgrounds but high target area consistency (as in the Big Ben image category), we obtain superior performance using the new sketch formation technique. To exhibit the high efficiency of the proposed algorithm for SBIR, we choose several categories in Flickr15k to display our retrieval results in Fig. 9.

The query sketches are in the black dotted rectangle box, and the top 5 retrieval image results and the corresponding generated sketches are in the red rectangles. From left to right, the sketches are obtained from the categories of duck, starfish, Sydney Opera House, Taj Mahal and Eiffel Tower.

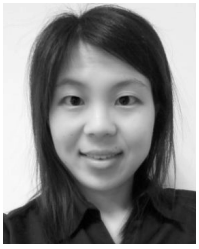
From Fig. 9, the retrieved images together with the generated sketches exhibit high semantic correspondence with the related retrieval sketch. There could exist inaccurate retrieval results in other categories; mistaken images might be retrieved due to the produced sketch having a similar form to the query sketch. In the step of feature matching, these images will be arranged in front and subsequently retrieved. Although inaccurate results could occur, most of the retrieval results are closely related to the query sketch in our work.

V. CONCLUSIONS

Sketch generation from natural images is an essential part of SBIR. In this paper, we employ a saliency detection method to generate clean and concise sketches from photo images. This method, which is based on the content of the improved absorbing Markov chain, is applied for SBIR for the first time. The results show that the novel method can produce a concise and clean pseudo-sketch from a single image. In addition, this technique obtains better MAP retrieval results and shows high efficiency for SBIR.

REFERENCES

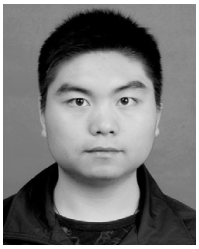
[1] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch objects?" *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–10, Jul. 2012.
[2] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *Int. J. Comput. Vis.*, vol. 106, no. 1, pp. 9–30, 2014.
[3] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "Transductive face sketch-photo synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 9, pp. 1364–1376, Sep. 2013.
[4] N. Wang, X. Gao, L. Sun, and J. Li, "Bayesian face sketch synthesis," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1264–1274, Mar. 2017.
[5] R. Hu and J. Collomosse, "A performance evaluation of gradient field HOG descriptor for sketch based image retrieval," *Comput. Vis. Image Understand.*, vol. 117, no. 7, pp. 790–806, Jul. 2013.
[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Jun. 2005, pp. 886–893.
[7] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 11, pp. 1624–1636, Nov. 2011.
[8] K. Chatfield, J. Philbin, and A. Zisserman, "Efficient retrieval of deformable shape classes using local self-similarities," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops (ICCV)*, Kyoto, Japan, Sep. 2009, pp. 264–271.
[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[10] J. M. Saavedra and B. Bustos, "An improved histogram of edge local orientations for sketch-based image retrieval," in *Proc. 32nd DAGM Conf. Pattern Recognit.*, Darmstadt, Germany, 2010, pp. 432–441.
[11] J. M. Saavedra, "Sketch based image retrieval using a soft computation of the histogram of edge local orientations (S-HELO)," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 2998–3002.
[12] R. Hu, M. Barnard, and J. Collomosse, "Gradient field descriptor for sketch based retrieval and localization," in *Proc. IEEE Int. Conf. Image Process.*, Hong Kong, Sep. 2010, pp. 1025–1028.
[13] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.
[14] G. Mori, S. Belongie, and J. Malik, "Efficient shape matching using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1832–1837, Nov. 2005.
[15] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "An evaluation of descriptors for large-scale image retrieval from sketched feature lines," *Comput. Graph.*, vol. 34, no. 5, pp. 482–498, Oct. 2010.
[16] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
[17] Q. Zhu, G. Song, and J. Shi, "Untangling cycles for contour grouping," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1–8.
[18] R. Kennedy, J. Gallier, and J. Shi, "Contour cut: Identifying salient contours in images by solving a Hermitian eigenvalue problem," in *Proc. CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 2065–2072.
[19] S. Marvaniya, S. Bhattacharjee, V. Manickavasagam, and A. Mittal, "Drawing an automatic sketch of deformable objects using only a few images," in *Proc. 12th Int. Conf. Comput. Vis.*, Florence, Italy, 2012, pp. 63–72.
[20] C.-E. Guo, S.-C. Zhu, and Y. N. Wu, "Primal sketch: Integrating structure and texture," *Comput. Vis. Image Understand.*, vol. 106, no. 1, pp. 5–19, 2007.
[21] Y. Qi, J. Guo, Y.-Z. Song, T. Xiang, H. Zhang, and Z.-H. Tan, "Im2Sketch: Sketch generation by unconflicted perceptual grouping," *Neurocomputing*, vol. 165, pp. 338–349, Oct. 2015.
[22] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[23] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3232–3242, Dec. 2010.
[24] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York, NY, USA: Wiley, 1971.
[25] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, Apr. 1970.
[26] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing Markov chain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1665–1672.
[27] J. Sun, H. Lu, and X. Liu, "Saliency region detection based on Markov absorption probabilities," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1639–1649, May 2015.
[28] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 2976–2983.
[29] Y. Kang, H. Lei, S. Lin, and Y. Li, "Sketch-based shape retrieval using orientation histogram with Gabor filters," in *Proc. 5th Int. Conf. Digit. Home*, Guangzhou, China, Nov. 2014, pp. 277–281.
[30] C. Adak, "Gabor filter and rough clustering based edge detection," in *Proc. Int. Conf. Human Comput. Interactions (ICHCI)*, Chennai, India, Aug. 2013, pp. 1–5.
[31] A. C. Cruz, B. Bhanu, and N. S. Thakoor, "Background suppressing Gabor energy filtering," *Pattern Recognit. Lett.*, vol. 52, pp. 40–47, Jan. 2015.
[32] X. Yan and X. Liu, "The improved two-dimensional Gabor filter based interest objects detection," in *Proc. 8th Int. Congr. Image Signal Process. (CISP)*, Shenyang, China, Oct. 2015, pp. 303–307.



XIANLIN ZHANG received the B.E. degree from Qufu Normal University and the master's degree from the Institute of Information Engineering, Northeastern University. She is currently pursuing the Ph.D. degree with the Beijing University of Posts and Telecommunications. Her research interests include sketch-based image retrieval, image pattern recognition, and deep learning.

XUEMING LI received the B.E. degree in electronics engineering from the University of Science and Technology of China in 1992 and the Ph.D. degree in electronics engineering from the Beijing University of Posts and Telecommunications (BUPT) in 1997. From 1997 to 1999, he was a Post-Doctoral Researcher with the Institute of Information Science, Beijing Jiaotong University. He has been with BUPT since 1999. In 2002, he was a Guest Lecturer with Karlsruhe University, Germany. His current research interests include digital image processing, video coding, and multimedia telecommunication. To date, he has undertaken many state and enterprise research and development projects, and he has authored or co-authored three books and over 50 papers in the field of multimedia information processing and transmission. He is a Senior Member of the Chinese Institute of Electrics and the China Society of Image and Graphics.

SHUXIN OUYANG received the B.E. degree from the Beijing University of Posts and Telecommunications, where she is currently pursuing the Ph.D. degree. She is also a Visiting Ph.D. Student with the Queen Mary University of London. Her research interests include cross-modality face matching, heterogeneous face recognition, and interactive learning.



YANG LIU received the bachelor's degree from Hunan University in 2010 with a major in communication engineering and the Ph.D. degree in engineering and information processing from the Beijing University of Posts and Telecommunications in 2015. He is currently a Lecturer with the School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications. His main research directions are machine learning, computer vision, and image processing.